

Open Data and Water Board Data Management

July 10, 2018 Board Meeting – Item 5

Greg Gearheart, Director

Office of Information Management and Analysis

State Water Resources Control Board / CalEPA

Water Board Data

The State Water Board collects data to help manage:

- water quality
- drinking water safety and
- water availability for 40 million Californians.

The data the Board collects inform almost every aspect of Californian's lives, are critical to the health of California citizens, and are vital in the state's economy.

In addition to the human uses, water data we manage are vital in managing the environmental needs of the state

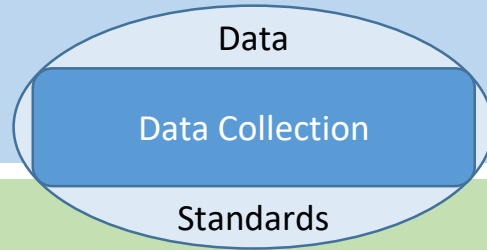
Why do we collect, store and use data?

- 1.To inform our **data-driven management** and planning activities – performance report cards, workplans, resource assignment, evaluating program effectiveness, and many others examples;
- 2.To inform our **critical decisions** regarding our mission(s) and water management responsibilities – water allocation and use, water quality planning and “policies,” permitting, program prioritization, etc.; and
- 3.To provide **transparency** to our many partners and stakeholders for their use, interests and purposes.

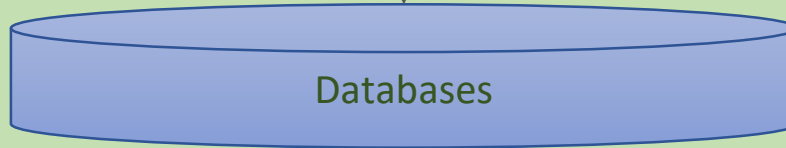
How do we currently manage our data?

- We **collect** some of it from our stakeholders, some we enter ourselves, some we acquire from other sources
- We **store** some of it on our own servers, some on the cloud, some is stored on desktop and other personal computers
- We **make it accessible** various ways, including public search forms and open data publishing
- We **use** it by transforming it into information via software tools like spreadsheet, GIS and business intelligence

Stage 1 - Collecting Data



Data
Checkers



S2 - Storing Data

S3 - Publishing Open Data

CalData data.ca.gov Portal

Water
Rights
Dataset
s

Water
Quality
Datasets

Web
Services

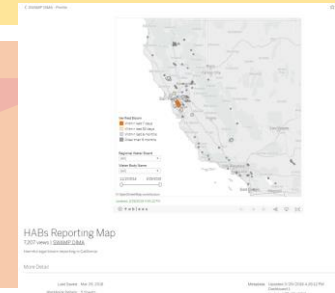
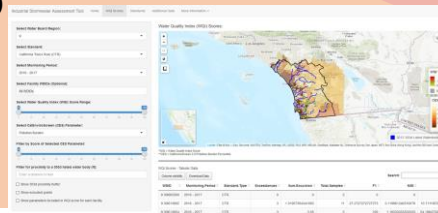
Other Open Data

Federal

CA DWR

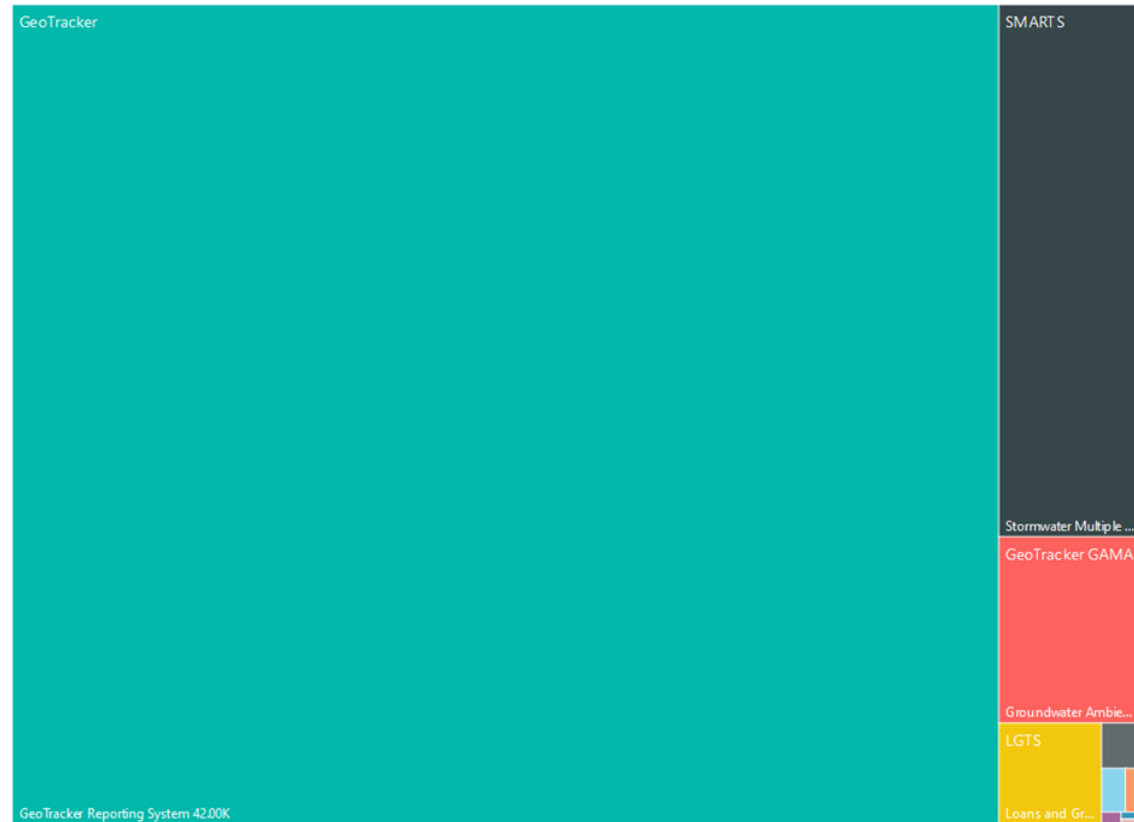
APIs / Web
Services

S4 - Turning Data into Information

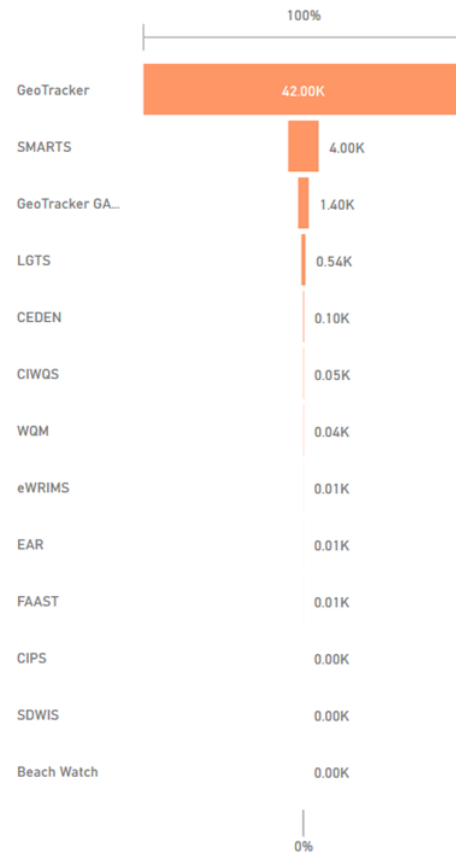


total data stored (GB) and First general description of contents by database and full name

database GeoTracker SMARTS GeoTracker GAMA LGTS CEDEN CIWQS WQM eWRIMS EAR FFAST CIPS SDWIS Beach Watch



total data stored (GB) and First general description of contents ...



The Open and Transparent Water Data Act (AB1755)

AB1755 requirements operative to WBs

- DWR is leading effort to build platform, develop protocols, etc.
- Water Board and WQ Monitoring Council is supporting DWR effort(s)
- Water Board must publish open data by September 2019. In addition, by August 2020 we must:
 - 12415. The statewide integrated water data platform created pursuant to Section 12410 shall, at a minimum, do all of the following:
 - (a) Integrate existing water and ecological data information from multiple autonomous databases managed by federal, state, and local agencies and academia using consistent and standardized formats.
 - (b) Integrate the following datasets, as available:
 - (1) The department's information on State Water Project reservoir operations, groundwater use, groundwater levels, urban water use, and land use.
 - (2) The state board's data on water rights, water diversions, and water quality through California Environmental Data Exchange Network (CEDEN)

What is “Open Data?”

- Open data is a term used around the world for past 6+ years to describe specific test of data accessibility and interoperability
- Publicly available
- Machine readable
 - Two dimensional flat files and/or available via APIs and/or web services
 - Curated and structured in a manner that makes it accessible and useful
- Well documented
 - Metadata – a robust and structured guide to the data to help users find context, limitations, issues including data dictionary

5 Guiding Principles for WB Data Management

1. **Make Data Accessible (“Open First”)**: our organization values transparency and strives to make all critical data available in machine readable datasets with metadata and data dictionaries
2. **Understand Data Quality and Integrity**: our data are thoughtfully planned, of known and useful quality, and we deploy practices to protect its integrity with standards and protocols
3. **Improve Data Literacy**: our whole organization understands its data needs and responsibilities, can speak the language of data science the staff and managers have robust data science capacity

5 Guiding Principles for WB Data Management (cont.)

4. **Use Data to Govern:** our organization uses data to govern, or makes decisions that are in the best interest of our mission(s)
5. **Govern our Data:** our organization takes proactive steps to develop effective data and information technology management practices to ensure our data flows to where it is needed in a timely manner while complying with our data sharing policies

Data Management Strategic Elements

1. Open Data Strategy
2. Data Literacy Strategy
3. Data Driven Management Strategy
4. Quality Management Strategy
5. Data Governance and Administration Strategy

Open Data Strategy

1. Civic engagement projects
 - a. Water Data Challenges, etc.
2. Make “flat file” pipelines of raw, high value data from all our priority databases applications
3. Identify data stewards in every part of the organization with charge and capacity to curate datasets around org
4. Datasets published on open data platforms, machine readable, refreshed as often as “needed” and well documented
5. Services and tools (and training) to support needs to analyze and visualize data (e.g., Tableau, etc.)
6. Predictive analytics, machine learning, fun stuff!

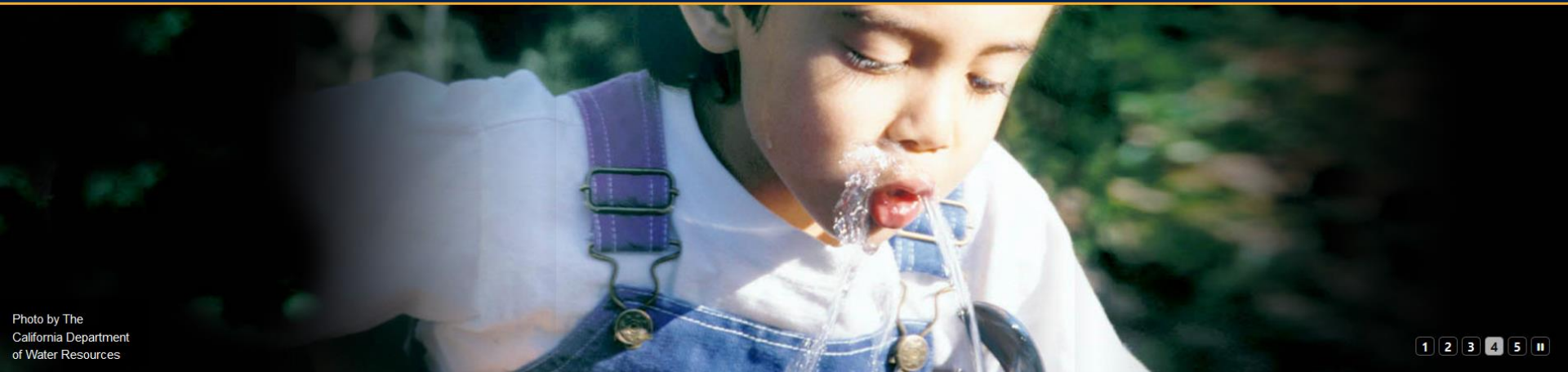


Photo by The
California Department
of Water Resources

1 2 3 4 5 ||



2018 Safe Drinking Water Data Challenge

2018 California Safe Drinking Water Data Challenge



Imagine H2O will award \$1,500 and select accelerator services to a team that best demonstrates Impact, Scalability, Innovation in their solution. [More information and terms](#)

Data Literacy Strategy

1. Training and capacity development
 - a. Organization (culture, systems and access to business intelligence tools)
 - b. Individual (staff vs. managers vs. executives)
 - c. Blend of customized online, in person and in-house providers
2. Engagement (internal and external)
 - a. Data concepts and terms in regular work products and conversations
 - b. Data science club (for projects, sharing, etc.)
 - c. Targeted projects to practice what is learned

Data Driven Management Strategy

1. Migrate performance report cards to open data-based visualizations and interactive tools
 - a. Program tools
 - b. Executive data storytelling (phased in over the next year)
 - i. Critical path stuff
 1. Business intelligence software distributed to staff, managers and executives
 2. A storytelling platform
 3. open data web services that work reliably

Data Governance and Administration

1. Governance - to address questions like:
 - a. Who reviews and approves the open data datasets?
 - b. What new data does the org need to do its job better?
2. Coordination with data collection and storage methods and approaches
 - a. How we collect and store data to make it easier to flow towards flat files, open data, etc.
 - b. When to use “spreadsheets” vs. building new DBs, etc.
3. Legal and privacy compliant practices in place

Next Steps

1. OIMA to conduct a strategic review of database governance and delivery for our primary, enterprise database applications. This includes an inventory and assessment of open data readiness for all existing datasets at the Water Boards.
2. Data Management Strategic Plan with 5 Elements and Inventory (as described above) by March 1, 2019
3. Quarterly progress reports to the Board and alignment of reporting with our performance report

Data Driven Management Examples

Water Quality Status Reports (2017 and 2018):

- https://www.waterboards.ca.gov/resources/data_databases/wq_status_report.html

FY2016/17 Water Boards Performance Report Story:

- <https://arcg.is/z5Km1>